

Lecture 9 presents details of structure-structure alignments and various web-based servers for their calculation. Alignments are important for structural annotation of new crystal structures. A typical query is - Is the newly determined structure unique or simply a structure of a well-recognized protein fold? Moreover, these methods are important for protein structure predictions.

Let us run through an exercise to anchor our understanding of structural alignments and the mechanics of their computation.

Topic 1

Go to rcsb.org and search the PDB for a crystallographic structure that is not currently annotated in SCOP, CATH or Dali (or other fold library). Perhaps check the newly released entries in PDB for a structure or implement an alternative search of PDB. Use structure alignment methods to propose a fold type for your selected structure. For calculation of the alignments, you will most likely have to download the structure to your computer (namely, Desktop) and then upload the structure to a web-based server for a comparison with known structures (a less time consuming option is to use a PDB subset of non-redundant topological folds).

Use Dali, FATCAT and or other methods (see, e.g., http://en.wikipedia.org/wiki/Structural_alignment_software) that allow the uploading of a structure to their server. Do not use sequence alignment techniques to search for the fold classification. The focus is on protein folds and the search of structural databases for 3D comparisons.

Because there are many available structures to explore, try not to duplicate the work of a posted result by a classmate. Keep in mind, however, alignment calculations may take some compute time, so use your time judiciously.

Report the following:

Describe your selected protein in terms of its functional annotation (if known), resolution and the number of residues. A suggested fold type and the RMSD values calculated from the alignments of your protein with other proteins and their sequence identities – namely, the output from Dali, FATCAT, etc. Note that there are many different computational schemes for calculating structure-structure alignments of proteins and their scoring methods (see, e.g., http://en.wikipedia.org/wiki/Structural_alignment). A molecular-graphics image of an alignment between your selected protein and the calculated homologue or structure neighbor (the so-called remote homologue).

Crystallographic structure: 6N7Z

6N7Z has a residue count of 375 and only a single unique protein chain. It is a thienopyrimidine-based allosteric inhibitor of human farnesyl pyrophosphate synthase and is classified as a transferase/transferase inhibitor.

Structure Alignment methods:

Download 6N7Z pdb file to upload to DALI for structural alignment against any other known protein structures. The top two results of structural neighbors are 6ely-B and 2vlc-A. As seen later on in FATCAT, 6ely-B is not structurally similar when examined in a pairwise alignment while 2vlc-A is. Hence forth, I shall only be comparing 6N7Z to 2vlc-A. According to the DALI alignment, 2vlc-A has a rmsd of 1.3 with a 67% sequence identity.

	No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
<input type="checkbox"/>	1:	6ely-B	39.2	1.3	260	264	61	PDB	MOLECULE: MISTLETOE LECTIN I;
<input type="checkbox"/>	2:	2vlc-A	36.1	1.3	262	518	67	PDB	MOLECULE: TYPE 2 RIBOSOME-INACTIVATING PROTEIN CINNAMOMIN

```

      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16     17     18     19     20     21     22     23     24     25     26     27     28     29     30     31     32     33     34     35     36     37     38     39     40     41     42     43     44     45     46     47     48     49     50     51     52     53     54     55     56     57     58     59     60     61     62     63     64     65     66     67     68     69     70     71     72     73     74     75     76     77     78     79     80     81     82     83     84     85     86     87     88     89     90     91     92     93     94     95     96     97     98     99    100    101    102    103    104    105    106    107    108    109    110    111    112    113    114    115    116    117    118    119    120    121    122    123    124    125    126    127    128    129    130    131    132    133    134    135    136    137    138    139    140    141    142    143    144    145    146    147    148    149    150    151    152    153    154    155    156    157    158    159    160    161    162    163    164    165    166    167    168    169    170    171    172    173    174    175    176    177    178    179    180    181    182    183    184    185    186    187    188    189    190    191    192    193    194    195    196    197    198    199    200    201    202    203    204    205    206    207    208    209    210    211    212    213    214    215    216    217    218    219    220    221    222    223    224    225    226    227    228    229    230    231    232    233    234    235    236    237    238    239    240    241    242    243    244    245    246    247    248    249    250    251    252    253    254    255    256    257    258    259    260    261    262    263    264    265    266    267    268    269    270    271    272    273    274    275    276    277    278    279    280    281    282    283    284    285    286    287    288    289    290    291    292    293    294    295    296    297    298    299    300    301    302    303    304    305    306    307    308    309    310    311    312    313    314    315    316    317    318    319    320    321    322    323    324    325    326    327    328    329    330    331    332    333    334    335    336    337    338    339    340    341    342    343    344    345    346    347    348    349    350    351    352    353    354    355    356    357    358    359    360    361    362    363    364    365    366    367    368    369    370    371    372    373    374    375    376    377    378    379    380    381    382    383    384    385    386    387    388    389    390    391    392    393    394    395    396    397    398    399    400    401    402    403    404    405    406    407    408    409    410    411    412    413    414    415    416    417    418    419    420    421    422    423    424    425    426    427    428    429    430    431    432    433    434    435    436    437    438    439    440    441    442    443    444    445    446    447    448    449    450    451    452    453    454    455    456    457    458    459    460    461    462    463    464    465    466    467    468    469    470    471    472    473    474    475    476    477    478    479    480    481    482    483    484    485    486    487    488    489    490    491    492    493    494    495    496    497    498    499    500    501    502    503    504    505    506    507    508    509    510    511    512    513    514    515    516    517    518    519    520    521    522    523    524    525    526    527    528    529    530    531    532    533    534    535    536    537    538    539    540    541    542    543    544    545    546    547    548    549    550    551    552    553    554    555    556    557    558    559    560    561    562    563    564    565    566    567    568    569    570    571    572    573    574    575    576    577    578    579    580    581    582    583    584    585    586    587    588    589    590    591    592    593    594    595    596    597    598    599    600    601    602    603    604    605    606    607    608    609    610    611    612    613    614    615    616    617    618    619    620    621    622    623    624    625    626    627    628    629    630    631    632    633    634    635    636    637    638    639    640    641    642    643    644    645    646    647    648    649    650    651    652    653    654    655    656    657    658    659    660    661    662    663    664    665    666    667    668    669    670    671    672    673    674    675    676    677    678    679    680    681    682    683    684    685    686    687    688    689    690    691    692    693    694    695    696    697    698    699    700    701    702    703    704    705    706    707    708    709    710    711    712    713    714    715    716    717    718    719    720    721    722    723    724    725    726    727    728    729    730    731    732    733    734    735    736    737    738    739    740    741    742    743    744    745    746    747    748    749    750    751    752    753    754    755    756    757    758    759    760    761    762    763    764    765    766    767    768    769    770    771    772    773    774    775    776    777    778    779    780    781    782    783    784    785    786    787    788    789    790    791    792    793    794    795    796    797    798    799    800    801    802    803    804    805    806    807    808    809    810    811    812    813    814    815    816    817    818    819    820    821    822    823    824    825    826    827    828    829    830    831    832    833    834    835    836    837    838    839    840    841    842    843    844    845    846    847    848    849    850    851    852    853    854    855    856    857    858    859    860    861    862    863    864    865    866    867    868    869    870    871    872    873    874    875    876    877    878    879    880    881    882    883    884    885    886    887    888    889    890    891    892    893    894    895    896    897    898    899    900    901    902    903    904    905    906    907    908    909    910    911    912    913    914    915    916    917    918    919    920    921    922    923    924    925    926    927    928    929    930    931    932    933    934    935    936    937    938    939    940    941    942    943    944    945    946    947    948    949    950    951    952    953    954    955    956    957    958    959    960    961    962    963    964    965    966    967    968    969    970    971    972    973    974    975    976    977    978    979    980    981    982    983    984    985    986    987    988    989    990    991    992    993    994    995    996    997    998    999   1000

```

Above is the Dali Alignment of the portion of 6N7Z that aligns with 2vlc-A. According to this alignment, the 6N7Z has relatively long regions of extended sheets interspersed with some coils and very few helices located right before position 500 of the alignment.

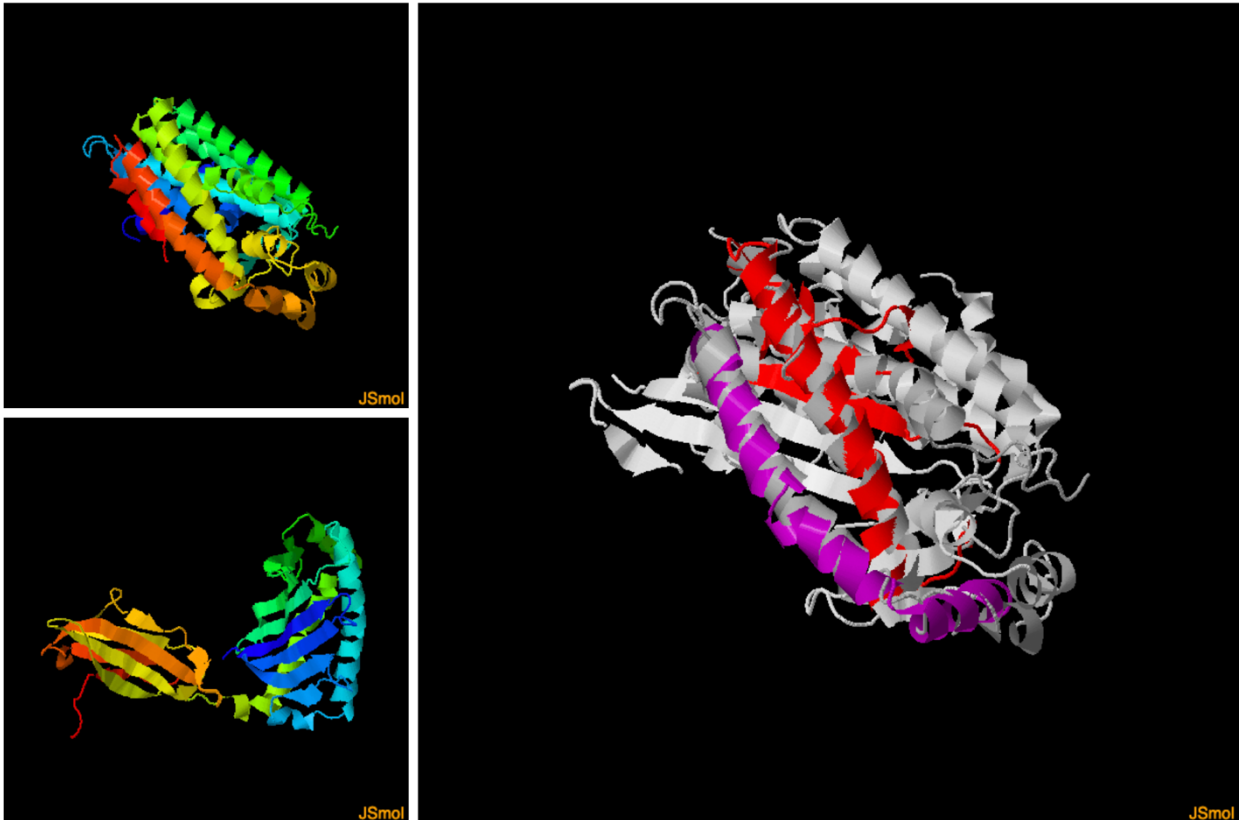
Go to RCSB and view structure similarities. No structure alignment is available for 6N7Z.F, but is represented by 4NUA.A which has > 95% sequence identity and is the crystallographic structure for the effects of Lysine 200 and Phenylalanine 239 on FPPS. Interestingly, although 4NUA.A has a >95% sequence identity, it did not appear in the structural alignments at all from DALI.

Use FATCAT to do a pairwise comparison between 6N7Z pdb file and 6ely-B and 2vlc-A. According to FATCAT pairwise structural alignment, 6ely-B is NOT significantly similar to 6N7Z with a p-value of 9.98e-01.

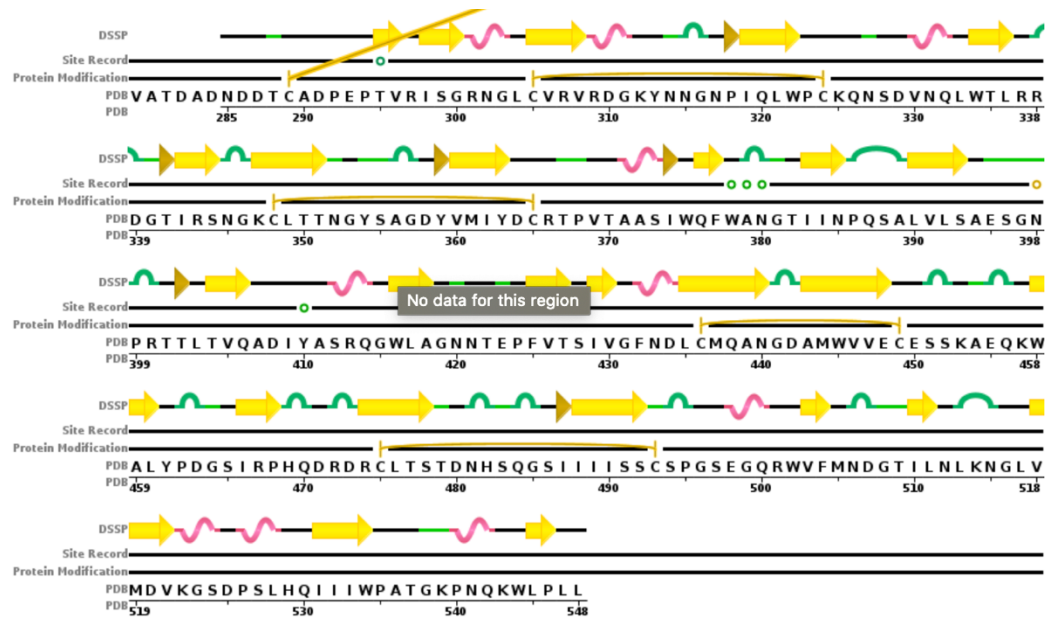
However, 2vcla-A IS significantly structurally similar to 6N7Z with a p-value of 4.75e-02. There are 97 equivalent positions with RMSD of 3.18 and 1 twist

JSmol display: str1 (left top); 2clva (left bottom); superposition (right), str1 in gray, 2clva in different colors (seperated by twists)

☒ Synchronize structure windows



Above is a superimposed image (right) of 6N7Z (top left) over the structurally similar 2vcla-A (bottom left).



Pictured above is the a sequence chain representation of the 2vlc-A sequence starting at around 280, where the overlap between 6N7Z occurs with 2vlc-A. Present in 2vlc-A are numerous beta strands and 3/10 helices. Since our highest structural alignment sequence ID was 2vlc-A from DALI and the pairwise structural alignment returned with high structural similarity from FATCAT, a good prediction of the fold type for 6N7Z can come from 2vlc-A. When searching SCOP for 2vlc-A, there were no proteins matching it, however, the entry of 2vcl-A is represented by 2VGR-C. Following 2VGR-C, I found that they are classified as a protein fold of: Ferredoxin-dependent bilin reductase-like. Extrapolating from this data, my best hypothesis is that 6N7Z is similar to 2vcl-A and can be considered with having a fold type of the same.

Topic 2

Is it scientifically meaningful to use structure alignments (via Dali, FATCAT, CE, etc) for protein functional annotation? Discuss the application of this approach and its limitations. Be specific.

As the 3D protein space is extremely challenging to predict and analyze due to the massive amounts of processing power needed to calculate all of the various 3D conformations of proteins, structural alignments for functional annotation are a great way to approximate a protein's 3D structure and extrapolate some hypothesis on functionality from said approximation. Of course, x-ray crystallography is going to be much better at getting the actual 3D structure of proteins, protein bioinformatic tools such as DALI, FATCAT, CE, etc. are all useful in getting a "close enough" answer for many practical needs. One way that I think of the usefulness of protein bioinformatics and other such heuristic tools is that they are useful in getting you in the right neighborhood, but maybe not the exactly correct house. However, it's easier to find the right house if you're somewhat nearby than if you're in a completely different city. While their usefulness in this era of protein bioinformatics cannot be understated, it is highly important that we recognize the shortcomings of using such

tools for protein analysis and annotation. Everything that these tools output are “best guesses” and should be treated as such. For important uses of certain proteins, further investigation and validation are required than just using these tools to find out their approximate structure and functionality, but lab experiments are necessary to go alongside the bioinformatic work to validate any results.

Due date is 18 Nov